

Proposta do Hackaton

ENSINO SUPERIOR

Escolha um dos 3 problemas propostos abaixo e esboce uma solução a partir dos dados disponibilizados.

Sobre os dados: São dados abertos disponíveis no Sistema Nacional de Informações Sobre Saneamento - SNIS. Estes dados são informados pelas empresas de saneamento de todo o Brasil, desde 1995, e são a base de decisão para políticas públicas e acompanhamento do desenvolvimento do setor de saneamento.

Dicas sobre a base de dados de forma geral:

- a) A base contém variáveis de diversos assuntos, muitos não relacionados ao problema que você escolher abordar. Selecione apenas as que podem ter alguma relação com seu problema.
- b) Nem tudo na base é confiável; lembre-se de [limpar](#) valores faltando, valores absurdos etc.
- c) A base vem toda "separada", cada linha se referindo a um prestador de um município em um ano específico. Agrupar ([1](#), [2](#)) (em ano e município ou em ano e prestador, por exemplo, a depender do problema) pode facilitar.

Problema 1: A perda de água na distribuição equivale a toda água que entrou para o sistema de abastecimento, mas que não chegou formalmente ao consumidor. No Brasil este indicador é de cerca de 40%. O Balanço Hídrico da International Water Association - IWA, classifica as perdas de água na distribuição em dois principais grupos: perdas reais que são os vazamentos nas redes de distribuição e unidades operacionais; e as perdas aparentes, que são aquelas que ocorrem por uso indevido da água ou submedição dos hidrômetros (devido ao desgaste natural do equipamento). Identificar onde estão as perdas e os custos envolvidos nas ações de combate às mesmas, estão no cerne das estratégias de redução de custos de todas as empresas de saneamento.

Desafio: buscar informações na base de dados que possam direcionar o combate a perdas de água. Que variáveis se correlacionam bem com as perdas de cada tipo?

Dicas: Entre todas as variáveis, é possível avaliar a importância delas por vários métodos de [Feature Selection](#), ou criando modelos de regressão interpretáveis como [modelos lineares](#) e [árvores de decisão](#).

Problema 2: As empresas de saneamento de abrangência regional, como CAGEPA, SABESP e COMPESA, embora possuam, praticamente, o mesmo modelo de negócio, apresentam resultados diferentes devido a estratégias de alocação de recursos e/ou maturidade tecnológica e de gestão. Entretanto, a distribuição das despesas parece seguir uma lógica em que o custo do consumo de energia elétrica e produtos químicos são bastantes expressivos. Identificar empresas que possuem melhores indicadores de custo/benefício desses insumos é uma

estratégia comum das empresas para identificar empresas que possuem melhores práticas de gestão e ou domínio tecnológico para ser seguido.

Desafio: avaliar quantitativamente os diferentes tipos de despesas de cada empresa. Buscar outliers e tentar explicá-los.

Dicas: [estatística descritiva](#) e técnicas de [clustering](#) podem ser úteis.

Problema 3: A universalização do saneamento é caracterizada por dois grandes problemas, levar água a 99% da população e esgotamento sanitário a 90% da população até 2033. Na Paraíba, assim como vários outros lugares do Brasil, a universalização do acesso à água potável está praticamente atingindo (cerca de 95% da população), já a cobertura do esgotamento sanitário ainda é baixa, sobretudo nos municípios de menor população. Portanto, diante da necessidade de investimento de grande monta, é necessário uma estratégia de priorização de investimentos que direcione os recursos para acelerar o processo de universalização.

Desafio: buscar insights na base de dados que possam ajudar a traçar estratégias para atingir a meta de universalização da rede de esgoto.

Dicas: algo no sentido de minimização de custos ou de tempo faria sentido, logo técnicas de [otimização](#) podem vir a ser úteis (se funciona ou não vai depender da criatividade ao escrever a função objetiva).